scientific reports

Check for updates

OPEN

Metagenomic profiling unveils the viral diversity in field-collected *Aedes* larvae from Central India employing nanopore sequencing

Ekta Gupta, Shashi Sharma, Paban Kumar Dash✉ & Manmohan Parida

Several arboviruses including Dengue, Chikungunya, Zika, West Nile and Japanese encephalitis viruses are emerging and re-emerging in many parts of the world over last two decades. Thus, environmental surveillance of the mosquito borne viruses employing latest next generation sequencing technology could enhance our comprehension about an impending outbreak, thereby, providing opportunity for timely intervention. In this study, *Aedes* larvae were collected from different locations of Gwalior, Central India, cultured and grown to adult and were screened utilizing metagenomic workflow in Oxford Nanopore platform. The results produced sufficient and valuable insights through demonstration of divergence of these mosquitoes' virome. Viral families associated with *Myoviridae*, *Mimiviridae*, *Iridoviridae*, *Bunyaviridae*, *Flaviviridae*, *Mesonivirdae* etc. were prevailing across the pools, varying in relative abundance. Viruses like *Betabaculovirus*, *Mimivirus*, Shamonda virus were reported in most pools in high abundance. Viral analysis leads to Phasi Charoen-like virus (PCLV), Nam Dinh virus (NDiV), Hubei mosquito virus (HMV), Wenzhou sobemo-like virus 4 (WSLV) findings across samples. This work reports the first successful metagenomic profiling of field-collected mosquitoes from Gwalior, Central India. Consequently, this technique might be employed to wide spectrum investigation of field mosquitoes, aiding in public health awareness about currently circulating viruses as a preparedness against future epidemic.

Keywords Nanopore, *Aedes*, Mosquito, Vector borne viruses, Viral metagenomics, Public health

Vector-borne infections are emerging as major health security concern, affecting human populations across continents. However, the effect is more pronounced in tropical and sub-tropical territories. Multiple factors including unplanned urbanization, an increase in human population density, increased contacts between humans and animals, climate change, etc, are primarily responsible for this startling rise in disease. The World Health Organization estimates more than 17% of all infectious diseases that affect humans are vector-borne infections which account for more than 700,000 deaths annually[1].

Among mosquitoes, genera *Aedes* and *Culex* are major vectors for transmission of a large no. of medically significant viruses. However, *Aedes aegypti* and *Aedes albopictus* are dominant vectors for transmission of viruses such as Dengue virus, Chikungunya virus and Zika virus[2]. Despite their importance in public health, there is a dearth of information regarding the ecology and epidemiology of viruses carried by mosquitoes that explains their host specificity, distribution, dominance, co-infection, and transmission[3]. India recorded resurgence of diseases such as Dengue, Chikungunya and Zika with annual outbreaks occurring across multiple states[4].

Traditional diagnostic approaches like culture dependent and contemporary real time reverse transcription PCR (qRT-PCR) techniques have proven to be useful in the detection of various known mosquito borne viruses[5–8]. The development of next generation sequencing (NGS) technologies has proved effectiveness in the exploration of mosquito microbial community. Techniques like 16s rRNA sequencing for bacterial profiling[9] and other shot gun metagenomics/ targeted metagenomic techniques[10], could successfully report and characterize mosquito borne viruses. However, these approaches are often expensive, time consuming and require expertise of bioinformatics. Third generation sequencing technologies like Nanopore sequencing offer a promising platform due to its portability, real time analysis feature and cost effectiveness[11]. The technology provides long reads through single-molecule chemistry, which also enables real-time or rapid analysis to save time. The portability helps in point of care or field-based sequencing as it could be performed on-site for better assessment of viruses,

Vector Biology Division, Defence Research and Development Establishment, Gwalior, Madhya Pradesh 474002, India. ✉email: pabandash.drde@gov.in

eliminating viral transportation cold chain issue. Sequencing depth and accuracy limitation is observed in nanopore sequencing technology. Nanopore sequencing is used by many researchers in exploring the mosquito virome and for mosquito surveillance programs[12].

In the rapidly developing discipline of metagenomics, genome from micro-organisms aid in learning more about their diversity, ecology and composition. This approach allows the identification of a wide variety of species found in different environmental and clinical samples, present in any microbial ecosystem. All of the viral sequences found in a sample, collectively termed as the virome, can be taxonomically classified to understand viral diversity and composition[13]. In the past few years, metagenomic sequencing has provided a substantial increase in exploration of novel viruses harboured by mosquitoes, revealing the vast complexity and extensive diversity of RNA and DNA viruses present in these mosquitoes. Virome analysis in mosquitoes is a preferred method of detecting the presence of viruses of interest, which are typically insect borne viruses (ISVs) or associated with human infections leading to potential public health significance. Virome research has the ability to identify viruses that could develop into future human pathogens as well as detect human infections that monitoring systems may have missed[14]. In a recent study, *Ae. aegypti* and *Cx. quinquefasciatus* mosquitoes have been found to have core viral communities that differ in diversity, where *Cx. quinquefasciatus* has greater viral diversity and variety of bacteriophages and *Ae. Aegypti* has a richer virome associated with eukaryotes[15].

Therefore, the goal of this study was to investigate the viral communities present in *Aedes* mosquitoes that were grown adult from larvae collected from various areas in Gwalior, Madhya Pradesh, as well as to assess the potential of metagenomic pipeline for broad spectrum vector surveillance that may contribute to improved monitoring and early detection of emerging arboviral threats that may eventually aid in public health.

Materials and Methods

Mosquito larvae collection, culturing and rearing

Mosquito larvae were collected from different areas of Gwalior (26° 13′ N, 78° 10′ E, 196 m altitude) between August to December 2023. The collection sites (Hazira, Civil Hospital, Shri Achleshwar Mahadev Temple, City Centre, Deen Dayal Nagar) were areas that were already reported with dengue infections (Fig. 1). Larvae were brought to Vector Biology Facility (temperature 28 °C; relative humidity 75–85% and 14:10 light: dark photo period) and were reared to adult (Fig. 2). Cotton pad soaked in 10% of sucrose solution was supplied to the adult mosquitoes. After emergence, adult mosquitoes were identified and separated on the basis of genus using standard morphological taxonomic keys. Only *Aedes* mosquitoes were selected for further downstream metagenomic analysis[16]. The male and female *Aedes* adult mosquitoes were segregated and pooled together in 1.5 ml centrifuge tube and were stored at − 80 °C until further processing[17].

Mosquito homogenization and RNA extraction

Five lab reared healthy mosquitoes were cold anaesthetized and spiked intrathoracically with 2 µl of previously lab grown West Nile virus (WNV) stock (10^7 PFU/ml titre) i.e. 2×10^4 PFU per mosquito, and pooled separately as a positive control for sequencing. Nine separate pools were prepared of adult female *Aedes* mosquitoes emerged from field collected larvae in number ranging from 40 to 120 depending on the sites and date of collection (Table 1). Total ten mosquito pools were prepared, homogenized in 2 ml tube with 500 µl of Dulbecco's Modified Eagle Medium (DMEM) and stainless-steel beads using Tissuelyser LT (Qiagen, Hilden, Germany) in cold condition for 15 min[18]. The homogenate was clarified by centrifuging at 4500xg for 10 min. Supernatant was filtered using 0.22 µm filter and collected in a new 1.5 ml microcentrifuge tube. The RNA was extracted from 140 µl of supernatant using the QIAamp viral RNA mini kit (Qiagen, Hilden, Germany). RNA was eluted in 50 µl of elution buffer. Five microlitre of eluted RNA was used as template for real time RT-PCR screening of West Nile Virus (WNV) employing SuperScript III Platinum One Step qRT-PCR kit (Invitrogen, MA, USA)[6]. The samples were analyzed in a 25 µl reaction volume with 12.5 µl of the 2x Master Mix, 0.5 µl each of forward primer and reverse primer (50 pmol/µl; 2 µM final concentration), 0.5 µl probe (10 pmol/µl; 0.4 µM final concentration), 0.5 µl enzyme mix, and 5.5 µl nuclease-free water. The thermal profile comprises of 50 °C for 20 min, 95 °C for 5 min, 40 cycles at 95 °C for 15 s and 60 °C for 45 s. The reaction was performed in 7500 Fast Real-Time PCR System (Applied Biosystems, MA, USA). No host RNA depletion or rRNA depletion step was performed. Field collected mosquito sample preparation involving pooling, homogenization and nucleic acid extraction was performed in biosafety level-2 (BSL-2) laboratory following institutional biosafety regulations. Intrathoracic inoculation with West Nile virus was conducted in biosafety level-3 (BSL-3) high containment facility following institutational and national biosafety guidelines.

Synthesis of double stranded cDNA and quantification

Double stranded cDNA was generated from above extracted RNAs using Maxima H Minus Double-Stranded cDNA synthesis kit (Thermo Fisher Scientific, MA, USA). For the purpose of double stranded cDNA synthesis, random hexamers were mixed with extracted RNA, which was then incubated at 65 °C for 5 min in a T1000 Thermal Cycler (Bio-Rad, CA, USA) and snap chilled[19]. Following this, a 4x first strand reaction buffer and a first strand enzyme mix were added. The mixture was incubated at 50 °C for 30 min followed by 85 °C for 5 min to stop the reaction. Subsequently, the second strand synthesis reaction was carried out by adding 5x second strand reaction buffer and second strand enzyme mix to the first strand product, which was then incubated at 16 °C for 60 min. The synthesized double stranded cDNA was quantified using Qubit 3.0 fluorometer and Qubit 1X dsDNA HS test kit (Thermo Fisher Scientific, MA, USA).

Metagenomic library preparation and nanopore sequencing

The synthesized double stranded cDNA was subjected to end repairing and dA-tailing using the NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs, MA, USA) according to manufacturer's instructions.

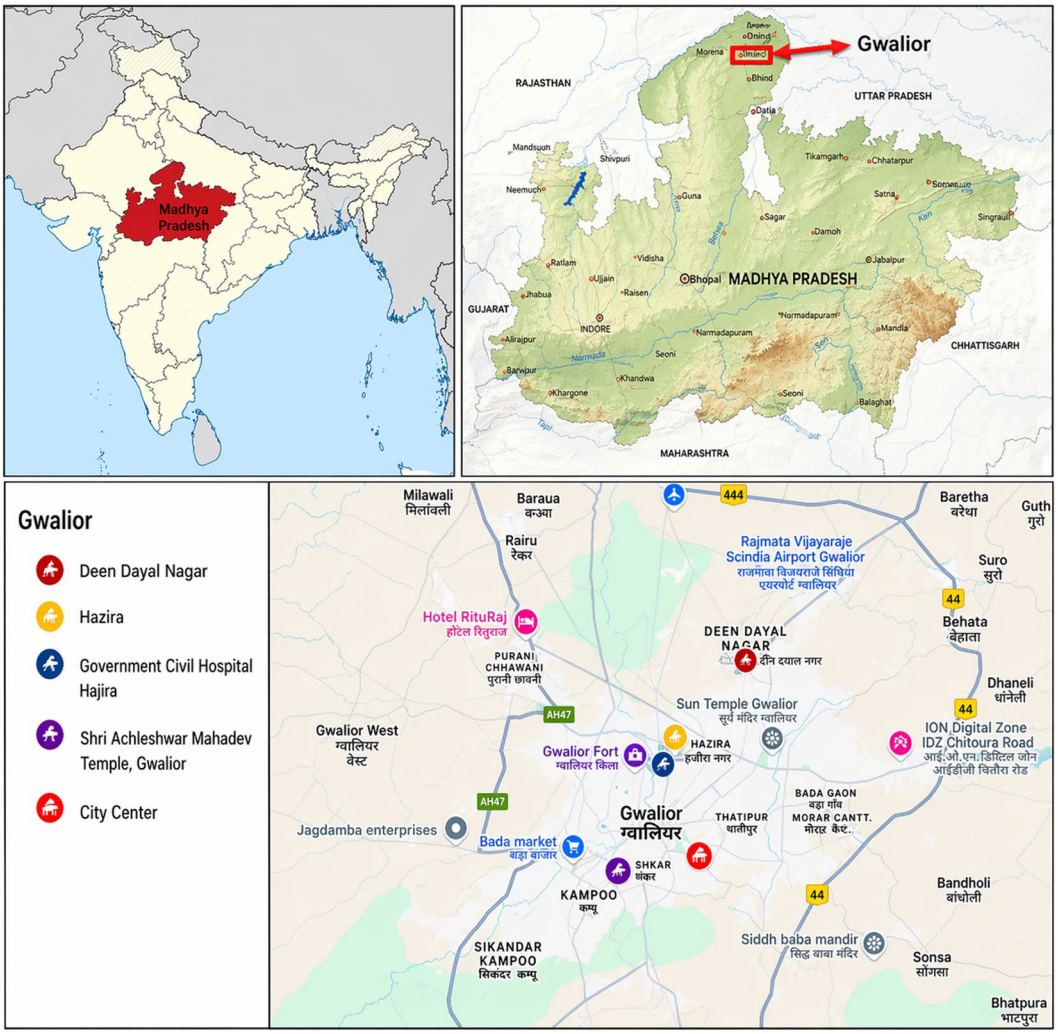

Fig. 1. Map of Gwalior as a geographical reference to show the locations from where mosquitoes were collected. Map of India represents Madhya Pradesh as a state in Central India. Gwalior Map reveals the collection sites of mosquitoes from locations like Deen Dayal Nagar, Hazira, Shri Achleshwar Mahadev Temple, Civil Hospital and City Centre. Maps created and sourced using Google Map (google.com/maps) and Maps of India (https://www.mapsofindia.com).

Native barcoding of the end repaired samples was done by employing Native Barcoding Expansion kit (EXP-NBD104, EXP-NBD114, Oxford Nanopore Technologies, Oxford, UK) followed by ligation sequencing protocol. All 9 samples were pooled and the pooled library was purified using 1X AMPure XP beads (Beckman Coulter, CA, USA). The final sample was ligated with sequencing adapters using Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies, Oxford, UK) with NEB Quick T4 DNA ligase (New England Biolabs, MA, USA) and Adapter Mix (AMX). The final library was purified using 0.5X AMPure XP Beads and Short Fragment Buffer (SFB) was used in washing steps. The library was eluted in 15 µl of Elution buffer[20]. The purified library was quantified using Qubit 3.0 fluorometer and Qubit 1X dsDNA HS assay kit (Thermo Fisher Scientific, MA, USA). The quality and pores of the nanopore flow cell (R9.4.1) was examined and primed according to manufacturer's instructions. The flow cell R9.4.1 was loaded with the prepared library, and a 72 h run was performed on a portable MinION Mk1B. MinKNOW software v21.11.7 executed base calling and real-time data analysis.

Computational analysis

Raw data was generated in the form of FAST5 format and was subjected to basecalling and demultiplexing using Guppy v5.1.13 from Oxford Nanopore Technologies[21]. Utilizing GUI based Commander 2.0 software (Genotypic Technology Pvt. Ltd., Bengaluru, India), the whole genome metagenome study examines and characterizes the species diversity and abundance in the samples. Adapter sequences were eliminated from raw reads using Porechop, low quality reads with q-scores < 8 and read lengths ≤ 100 bp were eliminated and statistical analysis of the reads were performed using NanoStat tool v1.6.0. The processed reads were aligned using the Centrifuge tool against the built-in centrifuge database, which included the NCBI-NR Database for Viral, Bacterial, Archaeal,

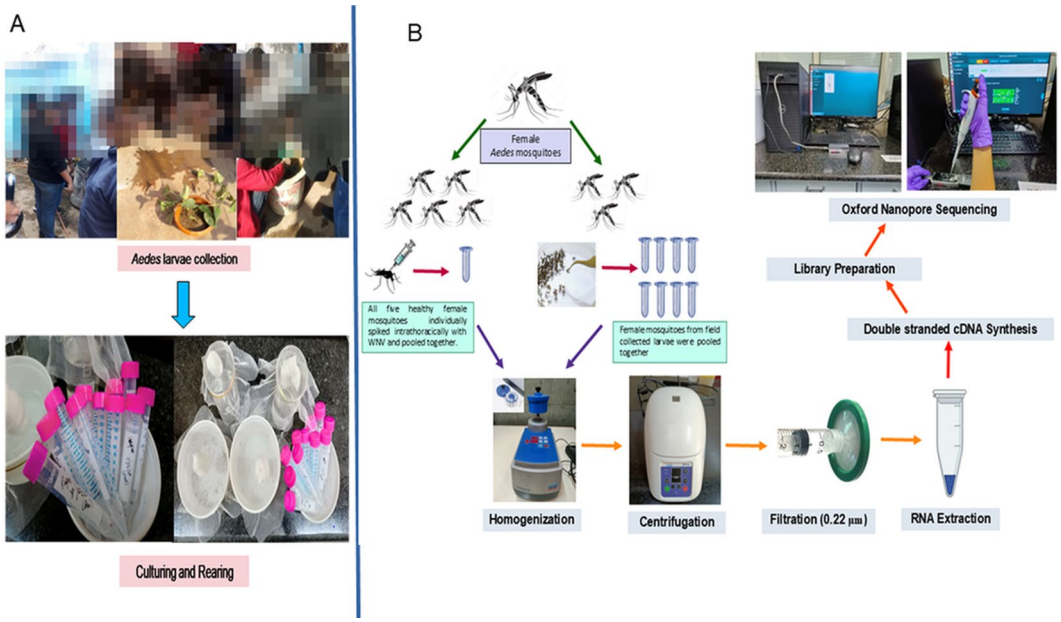

Fig. 2. Workflow illustrating mosquito larval collection, rearing, sample preparation, and metagenomic sequencing using Oxford Nanopore technology. (**A**) *Aedes* larvae collected from different areas of Gwalior and cultured. (**B**) Workflow comprising spiking of healthy mosquitoes with West Nile virus (WNV), pooling of spiked mosquitoes and field-collected mosquitoes for homogenization, centrifugation, filtration, RNA extraction, metagenomic library preparation and nanopore sequencing.

S.No.	Sample ID	Location	No. of Mosquitoes	Raw Reads	Processed Reads	Unclassified Reads	High Quality Data	Reads from Eukaryota, Archae, Bacteria	Viral Reads
01	DRDE/2023/01/PC	DRDE Lab	05	357,181	331,632	166,624	92.85	162,129	2879
02	DRDE/2023/01	Hazira	80	186,191	176,094	156,182	94.58	18,380	1532
03	DRDE/2023/02	Hazira	80	237,225	220,767	196,804	93.06	21,923	2040
04	DRDE/2023/03	Hazira	40	221,347	206,005	185,832	93.07	18,624	1549
05	DRDE/2023/04	DD Nagar	40	192,344	181,150	160,679	94.18	18,868	1603
06	DRDE/2023/05	DRDE	40	178,985	167,808	148,557	93.76	17,623	1628
07	DRDE/2023/06	Achleshwar	120	269,081	247,676	222,880	92.05	22,889	1907
08	DRDE/2023/07	Hazira	80	271,371	257,227	233,078	94.79	22,732	1417
09	DRDE/2023/08	DRDE	40	182,884	174,524	152,263	95.43	20,828	1433
10	DRDE/2023/09	Civil hospital	40	85,147	81,069	70,123	95.21	10,070	876

Table 1. Tabular data representing sample IDs, collection sites, no. of female mosquitoes pooled and metagenomic data generated. Sample ID: DRDE/2023/01/PC is the *in-vitro* WNV spiked sample as a positive control sample.

and Eukaryotic genomes for taxonomic classification. Centrifuge tool assigned individual reads to taxa based on their hits or sequence similarity and results were summarized from phylum to species level along with read counts. Results were generated in form of abundance plot, krona chart, kraken report, sankey plot, etc. (Fig. 3). Taxonomic classification reports generated by kraken tool was also visualized on an online tool Pavian version 1.2.0 (https://fbreitwieser.shinyapps.io/pavian/). For gathering more viral information, concatenated FASTQ

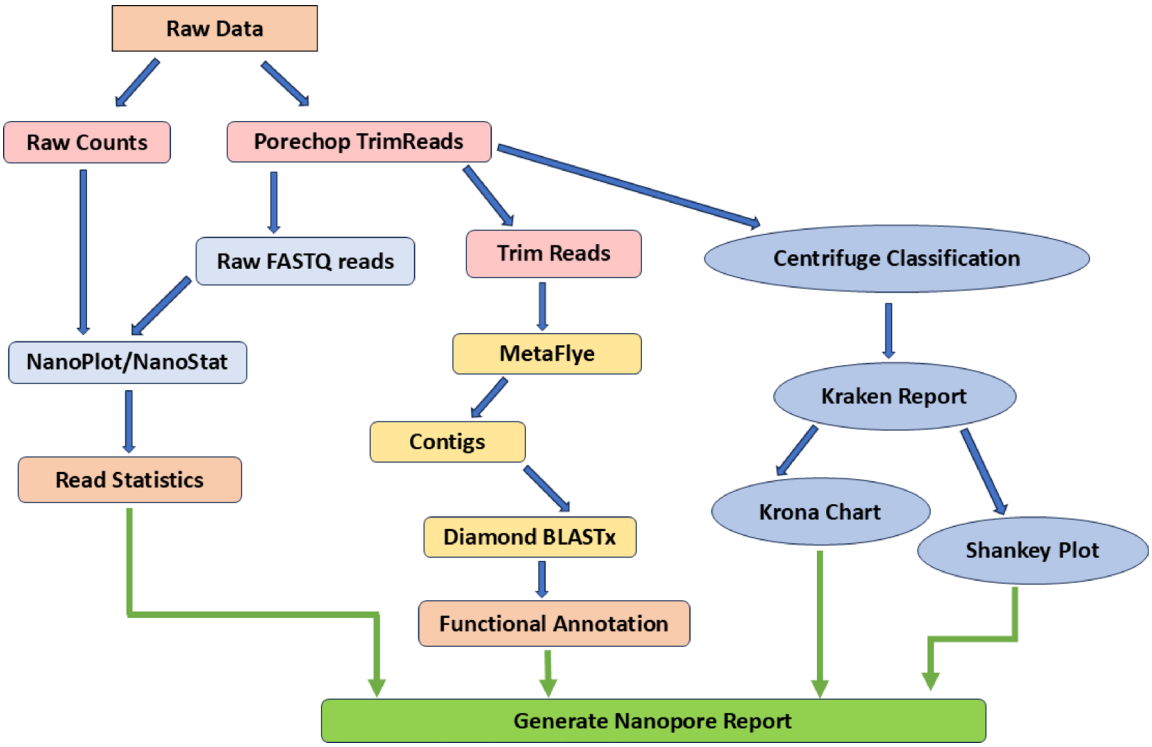

Fig. 3. Schematic diagram of computational analysis. Raw data generated from MinKNOW software is in FAST5 format which is further filtered and processed and aligned against Centrifuge database for taxonomic classification. The classification is reported in form of Kraken report, Krona chart and Sankey plot.

files of individual samples generated from Commander 2.0 software were run in Virus tool v2.14.5 in freely available web server Genome Detective Platform (www.genomedetective.com).

Phylogenetic analysis

NCBI Blast was performed on FASTA files and reference FASTA files were downloaded. Alignment of sequences was done using the MAFFT online server v7[22]. The aligned FASTA file was imported into MEGA v11, where the best-fit nucleotide substitution model was determined based on Bayesian Information Criterion (BIC). Tamura 3-parameter model with gamma distribution (T92 + G) was selected. Phylogenetic trees were constructed using the Maximum Likelihood method with 1000 bootstrap replications. Reference sequences were used based on NCBI BLAST nucleotide similarity (> 75–90%), query coverage and different geographic isolates. Details of reference sequences involving GenBank accession no., country name and collection year were also mentioned.

Results

Out of total number of 1210 *Aedes* larvae collected from different areas of Gwalior district and reared to adulthood, 560 were identified as females, and nine samples were prepared by pooling female *Aedes* mosquitoes together according to the collection sites.

Metagenomic analysis of West Nile virus spiked control sample (Sample ID: DRDE/2023/01/PC)

West Nile Virus (WNV) specific real time RT-PCR revealed that the virus spiked control sample (Sample ID: DRDE/2023/01/PC) was positive for WNV with cycle threshold (Ct) value of 24.9 whereas the nine field-collected mosquito test samples were tested negative for WNV. The positive control RNA reveals a Ct value of 18 (Figure S1). The metagenomic sequencing of the WNV spiked sample yielded 357181 raw reads, of which 331632 reads passed quality filtering. Nearly, 92.85% of total reads were retained as passed reads. Reads 162,129 obtained were mainly associated with Eukaryota (majorly *Homo sapiens*), Bacteria, Archae and viruses and 166624 unclassified reads (50%) (Fig. 4A and B). Krona report revealed viral reads associated with *Betabaculovirus* (88%), Shamonda virus (7%), *Flavivirus* (2%), *Caudovirales* (1%) and others including genera *Vesivirus* and *Tobamovirus*, and families *Phycodnaviridae*, *Retroviridae*, *Herpesviridae*, *Poxviridae*, *Mimiviridae and Iridoviridae*. Reads classified under the genus *Flavivirus* comprised of West Nile virus and Cell fusing agent virus (CFAV), thus confirming the presence of spiked virus WNV in the sample (Fig. 4C and D Sankey plot visualization further represented distribution and relative abundance of microbial taxa, with *Choristoneura occidentalis granulovirus* as the most abundant viral species (Fig. 5).

A

Sample ID	No. of Mosquitoes	Raw Reads	Processed Reads	Unclassified Reads	High Quality Reads	Reads from Eukaryota, Archae, Bacteria	Viral Reads
DRDE/2023/01/PC	05	357,181	331,632	166624	92.85	162,129	2879

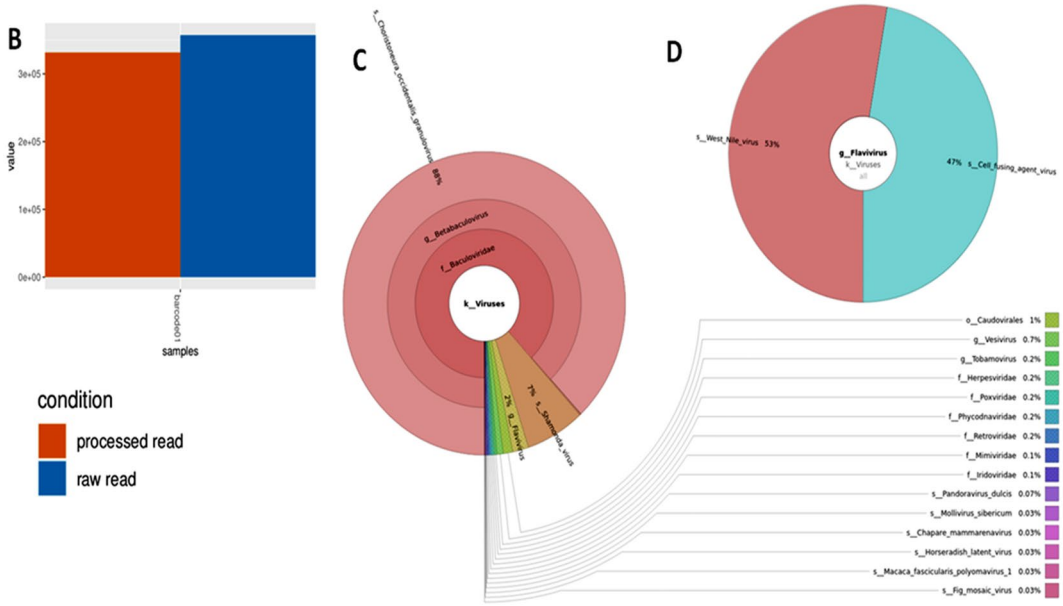

Fig. 4. (**A**) This tabular data showcases details from raw reads to processed reads, high quality data, reads from Eukaryotes, Archae, Bacteria and viral reads. (**B**) Read statistics plot represents raw and processed reads. (**C**) Krona chart of metagenomic data depicts viral reads and its association to different viral taxonomic families like 88% of *Betabaculovirus*, 7% of Shamonda virus, 2% of *Flavivirus*, 1% of Caudavirales and others included in less than 1% varying from 0.7 to 0.03% such as genera including *Vesivirus* and *Tobovirus* and families including *Phycodnaviridae*, *Retroviridae*, *Herpesviridae*, *Poxviridae*, *Mimiviridae*, *Irodoviridae*, etc. (**D**) Krona chart represents 2% of flavivirus data comprising West Nile virus (WNV) and Cell fusing agent virus (CFAV).

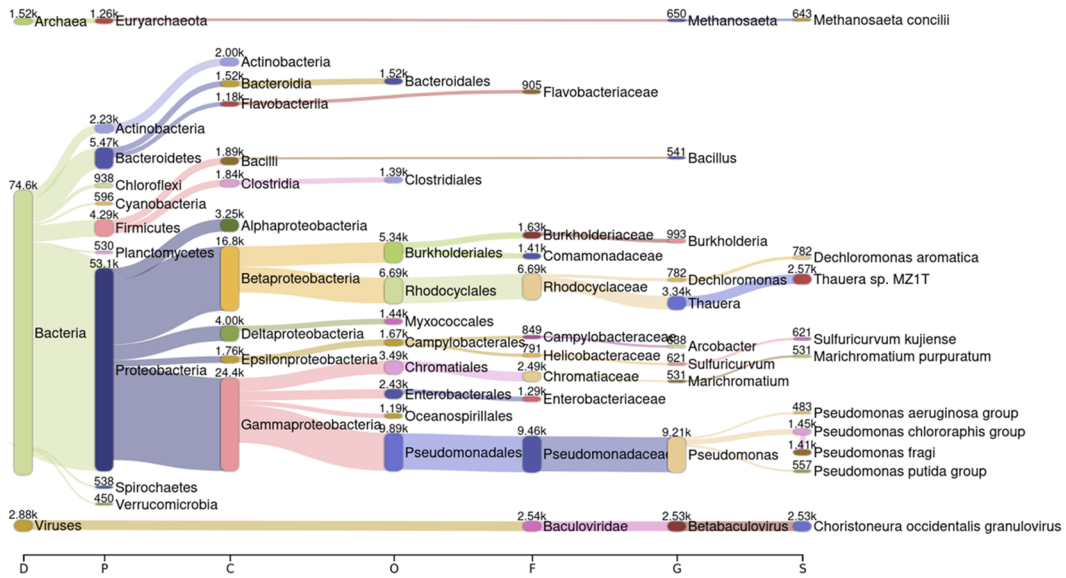

Fig. 5. Sankey diagram. It showcases metagenomic datasets revealing abundance of microbial community. The diagram illustrates hierarchy and reads distribution across taxonomic levels starting from Domain (D), Phylum (P), Class (C), Order (O), Family (F), Genus (G), to Species (S). Difference in width of flow suggests relative abundance of reads of assigned taxon. It describes the connection between different taxonomic ranks. The analysis indicated that *Choristoneura occidentalis granulovirus* of family *Baculoviridae* had the highest read count i.e. 2,533 reads.

Metagenomic profiling of field-collected mosquito pools

Nine mosquito pool samples were subjected to Oxford Nanopore sequencing by following the earlier mentioned pipeline for library preparation and metagenomic whole genome sequencing. Approximately, 4.29 million reads output was generated in 72 h run (Table 1). Taxonomy analysis in which reads were assigned to taxa based on their hits and results were summarized from phylum to species level along with read counts (Fig. 6; Figure S2, S3, S4; Table S1). The virome from these field collected mosquitoes was diverse comprising viral families like *Baculoviridae*, *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Mimiviridae*, *Poxviridae*, *Adenoviridae*, *Iridoviridae*, *Bunyaviridae*, *Flaviviridae*, *Mesoniviridae* etc. representing double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), and single-stranded RNA (ssRNA) viruses, including both negative-sense and positive-sense RNA viruses. Some virus taxa were found to be dominant across all samples including insect-associated baculoviruses like genera *Betabaculovirus* and *Alphabaculovirus* (family *Baculoviridae*). Bacteriophages-associated families like *Myoviridae* (genus *Tequatrovirus* or T4 virus) and Siphoviridae (genus *Tequintavirus* or C5 virus) were also widely distributed. Additionally, protozoa and vertebrates associated viruses such as genus *Mimivirus* (family *Mimiviridae*), genus *Mastadenovirus* (family *Adenoviridae*), genus *Iridovirus* (family *Iridoviridae*) and Shamonda virus (family *Peribunyaviridae*) were also detected in multiple samples thus demonstrating complexity and diversification of viral communities in mosquito populations (Figure S2).

Overview of insect-specific viruses (ISVs) in mosquitoes

NCBI BLAST of FASTA file derived after assembly of reads by Commander software revealed high abundance of insect-specific viruses, particularly Phasi Charoen-like virus (PCLV) and Humaita-Tubiacanga virus (HTV) across samples. For further viral characterization, concatenated FASTQ files obtained from Commander software were run on Genome Detective Platform. For meaningful interpretation, ISV presence was evaluated based on genome coverage, read count, read depth, nucleotide similarity and concordance. Reads with more than 50% nucleotide similarities, higher read count and genome coverage were mainly from viruses like *Phasivirus phasiense* or Phasi Charoen-like virus (PCLV; S, M and L segments), Nam Dinh virus (NDiV), Hubei mosquito virus 2 (HMV2), Hubei sobemo-like virus 41 (HSLV 41), Wenzhou sobemo-like virus 4 (WSLV4) as demonstrated in supplementary file (Figure S5). Viruses such as *Choristoneura fumiferana granulovirus* (family *Baculoviridae*), *Piry vesiculovirus* (family *Rhabdoviridae*) and viruses associated with vertebrates, plants, bacteriophages were detected with low nucleotide similarity, short alignment length, low read count (< 10 reads) and limited genome coverage (< 10%).

Phylogenetic analysis

Phylogenetic analysis was performed for insect-specific viruses identified in metagenomic datasets including Phasi Charoen-like virus (PCLV; S, M and L segments), Wenzhou sobemo-like virus 4 (WSLV-4), Humaita-Tubiacanga virus (HTV), and Nam Dinh virus (NDiV).

PCLV-S and L segment got clustered with Indian and Brazilian sequences suggesting strong evolutionary conservation and wide occurrence in distinct geographically located mosquito population, whereas PCLV-M

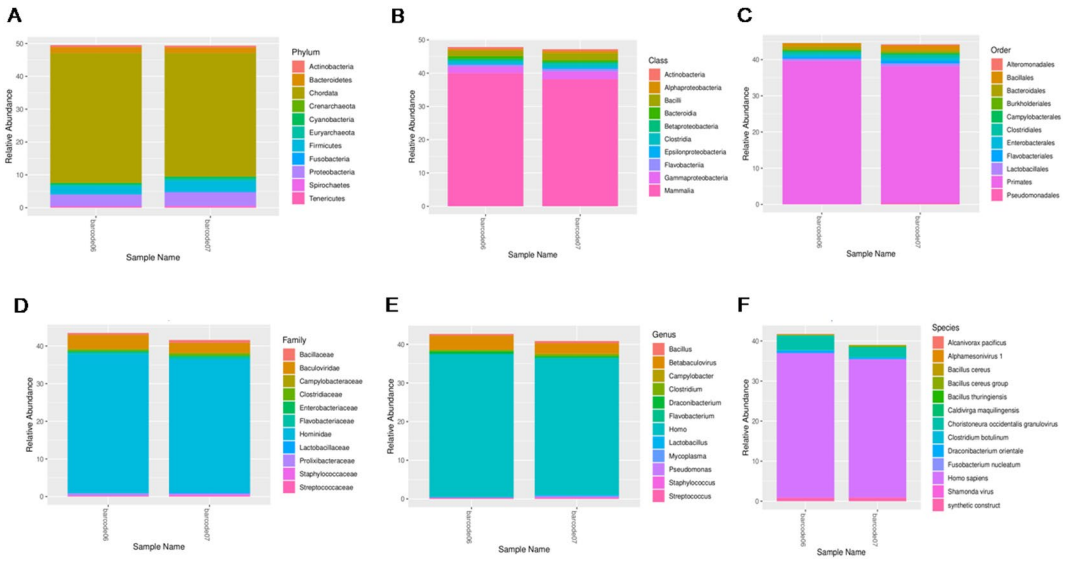

Fig. 6. Sample report illustrating hierarchical taxonomic classification (Phylum, Class, Order, Family, Genus and Species) and relative abundance of organisms, generated from centrifuge tool analysis. Graphical representation states taxonomic classification including Phylum, Class, Order, Family, Genus and Species with relative abundance of organisms on the basis of hits or reads of the sample. Each colour represent a particular taxon and proportion represents relative abundance of that taxon in sample. Sample IDs DRDE/2023/06 (Barcode 06) and DRDE/2023/07 (Barcode 07) are used for representation of analysis. Reports of all samples is mentioned in supplementary file (Figure S4).

segment got clustered with Indian, Brazilian and Kenyan sequences (Fig. 7A–C), illustrating slightly higher divergence as compared to PCLV-S and L. The clustering of present sample sequences supports its prevalence and genetic stability in *Aedes* mosquitoes.

Phylogenetic analysis of WSLV-4 sequences show grouping with previously reported sequences from China, Japan, Switzerland and Greece (Fig. 7D). The present sample sequences formed close related cluster within the clade, indicating high nucleotide similarity and close evolutionary relationships. The sequences show association with other sobemo-like viruses associated with mosquitoes such as Guangzhou sobemo-like virus and Nea chili luteo-like virus indicating close ancestry. Additionally, related clustering can also be seen with Guadeloupe mosquito virus with GenBank accession number and percent identity: MN053791.1 (91.09%), MN053799.1(86.4%), MW434816.1 (90.86%), OR670429.1(91.39%). The clustering pattern indicates that the sequences detected belong to WSLV-4.

Detected HTV sequences were also grouped with already reported global isolates from Guadeloupe, Brazil, Kenya and Trinidad and Tobago (Fig. 7E). This indicates the conserved relationship and confirmed that the sequence belongs to HTV. The clustering pattern demonstrates its wide presence in geographically distant mosquitoes.

Analysis of detected NDiV sequences were found to be clustered with reference isolates belonging to genus *Alphamesonivirus*, family *Mesoniviridae* (Fig. 7F). Close association of Indian sequences with previously reported sequences from China and Southeast Asia can be seen, thus confirming the taxonomic classification of NDiV and evolutionary relationship with mesoniviruses carried by *Aedes* mosquitoes across globe.

Identification of other viruses

The kraken report generated from each sample provides data that suggests presence of range of DNA and RNA viruses associated with insects, bacteria, vertebrate, protozoan and plants. Respective viral families were identified across all mosquito pools and number of reads assigned were also mentioned (Table S1). Among double stranded DNA viruses (dsDNA viruses), families such as *Baculoviridae*, bacteriophages related families like *Myoviridae*, *Siphoviridae*, *Podoviridae* and large DNA viruses belonging to families like *Mimiviridae* and *Herpesviridae* were detected frequently across all samples. RNA viruses (single stranded positive sense RNA virus (ssRNA (+)), single stranded negative sense RNA viruses (ssRNA (–)) and double stranded RNA viruses (ds RNA viruses)) comprised of families such as *Mesoniviridae* and *Flaviviridae* (insect-specific viruses), *Closteroviridae* and *Geminiviridae* (Plant viruses) and *Coronaviridae* and *Reoviridae* (vertebrate viruses) were present in most samples. Overall, these viruses show complex virome structure of *Aede* mosquitoes (Table 2).

Discussion

This study represents as first report on viral metagenomics of *Aedes* mosquitoes from Madhya Pradesh state in Central India. Unbiased nanopore oriented metagenomic analysis approach coupled with next generation sequencing (NGS) has transformed virus detection and surveillance over the last decade and has emerged as a useful tool for detection of various viral taxa without its prior information[23,24]. However, these methods are susceptible to restrictions related to sample type, sample quality, nucleic acid integrity, and sequencing depth, much like any other detection-based technology which can affect viral recovery and detection efficiency.

In this study, we did not incorporate host nucleic acid depletion or rRNA removal steps prior sequencing. This approach was adopted to retain unbiased nucleic acid in all samples and to evaluate performance of metagenomic pipeline in field-based setting. However, as expected, a high proportion of reads associated to host and bacteria were found which caused decrease in proportion of viral reads and also affected the genome coverage. The burden of host sequences or bacterial sequences may interfere with viral genome coverage resulting in less sensitivity in viral reads[25]. Many previously reported studies have shown advantage of introducing host depletion methods, rRNA removal, centrifugation/filtration steps, leading to enrichment of viral composition in sample, improved detection efficiency, sensitivity and coverage[26]. Nevertheless, metagenomic profiling pipeline in this study still provided detection of various important viruses under minimally processed conditions which can be further exploited for field-based detection. The findings in this study reveal the abundance of microbiome in mosquitoes and the higher potential of nanopore unbiased metagenomic sequencing over amplification and culture dependent techniques.

Since West Nile virus is less prevalent in *Aedes aegypti* than Chikungunya, Dengue, Zika in India, it was selected for intrathoracic spiking in mosquitoes. This was done to add a positive control in the experiment as well as to understand the inhibitory effect of a high titre virus to its background viral communities when compared with non-spiked mosquitoes. This experimental approach aided in evaluation of sensitivity, unbiased metagenomic pipeline and reliability of nanopore sequencing platform. Detection of insect specific viruses (ISVs) and other viral sequences in this study demonstrates the ability of random priming and reverse transcription used for cDNA synthesis and library preparation to detect wide range of RNA viruses in mosquito homogenates. This does not enrich viral sequences but aids in including viral sequences in sequencing library. ISVs identified from mosquito virome like Phasi Charoen-like virus (PCLV), Wenzhou sobemo-like virus 4 (WSLV4), and Hubei mosquito virus 2 (HMV2) etc. are also previously reported in mosquito populations[13,27,28]. In a recent study, PCLV and HTV has been found in abundance in *Ae. aegypti* mosquitoes[29]. These ISVs may impact mosquitoes' ability to transmit harmful viruses, even if they are unlikely to be linked to disease in vertebrates. The *Nam Dinh virus*, which is closely related to Cavally virus (CavV) and is member of the genus *Alphamesonivirus1* of the family *Mesoniviridae*, was originally identified from *Culex quinquefasciatus* in Shenzhen, China in 2011[30]. *Alphamesonivirus1* is the founding species of *Mesoniviridae* and is also widely found in mosquito virome. WSLV-4 was found in *Ae. albopictus* for the first time in China in 2013 but not much information was known at that time. Thus, a proper background study on ISV and its potent vector is required for more information[31]. A recent metagenomic study conducted in Guangzhou, China found WSLV-4 virus in their mosquito samples[32].

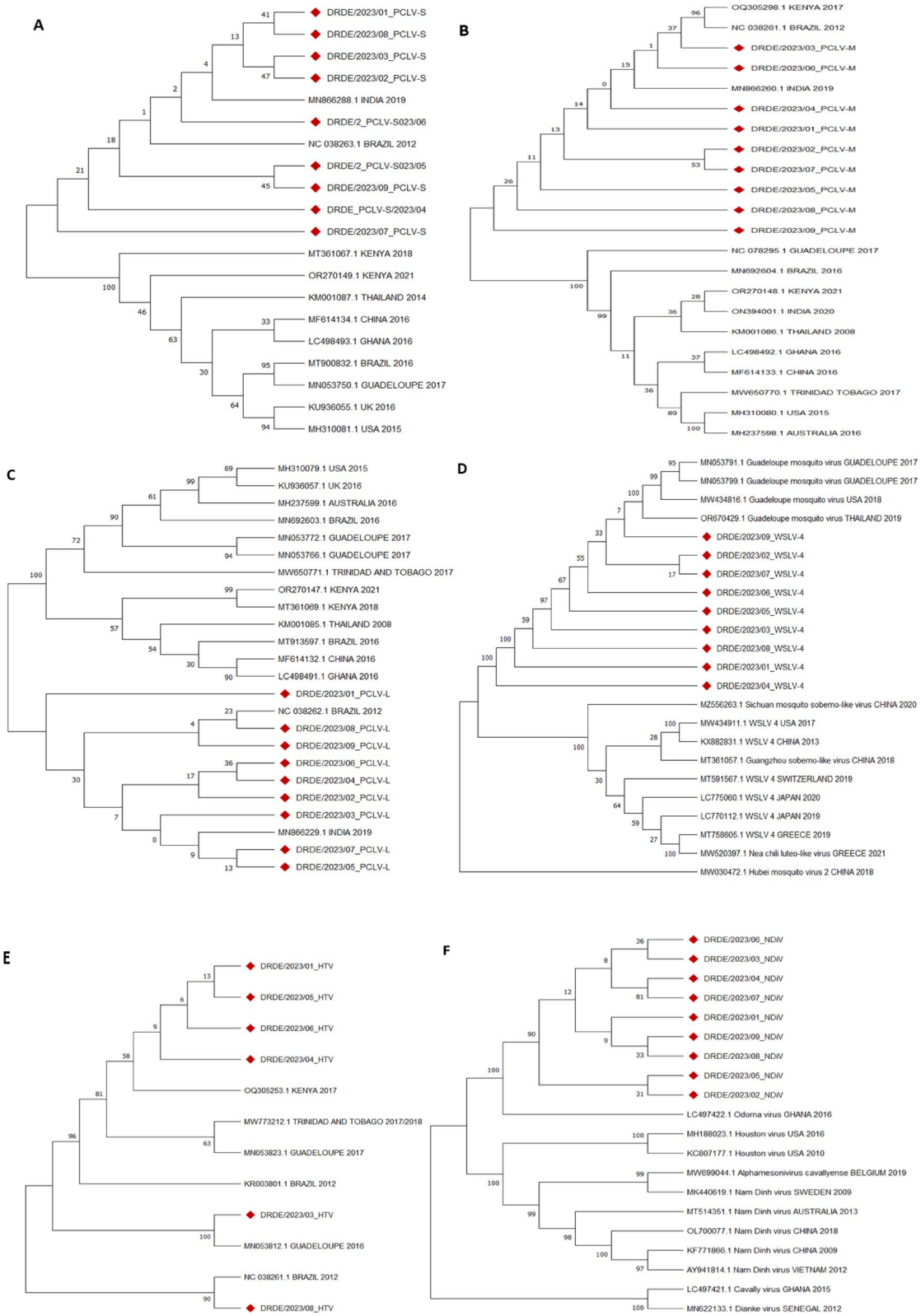

Fig. 7. Maximum likelihood phylogenetic analysis of viral sequences identified in *Aedes* mosquitoes. Phylogenetic tree was constructed based on nucleotide sequences of insect specific viruses aligned using MAFFT v7 online server. Phylogenetic analysis aid in determining evolutionary relationships of (**A**) Phasi Charoen-like virus – S segment (PCLV-S), (**B**) Phasi Charoen-like virus – M segment (PCLV-M), (**C**) Phasi Charoen-like virus – L segment (PCLV-L), (**D**) Wenzhou sobemo-like virus-4 (WSLV-4) (**E**) Humaita-Tubiacanga virus (HTV) (**F**) Nam Dinh virus (NDiV). Frequency values based on bootstrap replicates are indicated at branch nodes. Sequences of samples included in present study are highlighted using Sample ID and red diamond symbol.

Genome	Family	Associated viruses	Host range	Sample IDs involved
dsDNA viruses	*Baculoviridae*	*Choristoneura occidentalis granulovirus*, *Catopsilia pomona nucleopolyhedrovirus*	Insects (Lepidoptera, Diptera)	DRDE/2023/01, 02, 04, 05, 06, 07
	Myoviridae	Salmonella phage SSE121, Bacillus phage JBP901, Twortvirus, T4 virus	Bacteria	Across most samples
	Siphoviridae	Lactobacillus phage LLKu	Bacteria	Across all samples
	Podoviridae	Enterobacter phage EcP1, Vibrio phage VBP32, Enterobacteria phage Bp4	Bacteria	DRDE/2023/02, 03, 05, 06, 07
	Poxviridae	Sheepox virus, Swinepox virus, Canarypox virus, Skunkpox virus	Vertebrates, arthropods	Across most samples
	Adenoviridae	Bat mastadenovirus WIV13, Porcine adenovirus 3	Mammals, birds	DRDE/2023/01, 05, 06, 07, 08
	Herpesviridae	Human herpesvirus 3, Elephant herpesvirus 4, Ranid herpesvirus 2	Vertebrates	Across most samples
	Mimiviridae	*Acanthamoeba polyphaga mimivirus*, *Cafeteria roenbergensis virus*	Protozoa	Across all samples
	Other dsDNA families like *Polydnaviridae*, *Marseilleviridae*, *Iridoviridae*, *Phycodnaviridae*, *Nudiviridae*	Viruses not specified	Insects, algae, protozoa	Across most samples
ssRNA (–)	*Bunyaviridae*	Shamonda virus, Tomato spotted wilt virus, Iris yellow spot virus	Insects, plants	DRDE/2023/05, 06 and others
ssRNA (+)	*Mesoniviridae*	Nam Dinh virus, Cavally virus, Hana virus, Casuarina virus	Mosquitoes	Across all samples
	Flaviviridae	Cell fusing agent virus (CFAV)	Mosquitoes	DRDE/2023/04, 05, 06, 07, 09
	Closteroviridae	Cucurbit chlorotic yellows virus, Beet yellows virus	Plants	DRDE/2023/01, 06
	Coronaviridae	Porcine coronavirus HKU15, Rabbit coronavirus HKU14	Mammals	DRDE/2023/02, 07, 08
	Other ssRNA families like *Dicistroviridae*, *Secoviridae*, *Caliciviridae*, *Luteoviridae*, *Narnaviridae*	Viruses not specified	Insects, plants, fungi	Across most samples
dsRNA viruses	*Geminiviridae*	*Grapevine geminivirus A*, Tomato leaf curl Gujarat virus	Plants	DRDE/2023/01, 02, 03, 06
	Reoviridae	Rotavirus D, Fako virus	Vertebrates, insects	DRDE/2023/06, 08

Table 2. Tabular representation of viruses found in different samples in metagenomic profiling of Aedes mosquitoes. It mentions genome type (double stranded DNA viruses (dsDNA viruses), single stranded negative sense RNA viruses (ssRNA (–)), single stranded positive sense RNA viruses (ssRNA (+)) and double stranded RNA viruses(dsRNA viruses)), families detected in profiling analysis, Associated viruses belonging to respective familes, viruses connected to host and Sample IDs in which this viruses are detected.

The possible role of such diverse ISVs mainly lies in restricting the transmission of arboviruses in mosquitoes thus preventing disease spread as in lot of studies it has been found that presence of ISVs inhibits arboviral replication and hence decreases viral titre[33–35]. ISVs can act as a biological control for mosquitoes which can also be exploited to eradicate mosquito population from certain areas hence prevent mosquito borne diseases.

Viral sequences associated with vertebrate and plants were also detected. However, these findings require careful interpretations. The presence of such viral sequences may suggest detection of fragments of nucleic acid derived from blood meal, environmental interaction or similarity with viral regions in sequences. This cannot be interpreted as active infection in mosquitoes. Also, larvae collected and used in this study were reared to adulthood and hence the presence of these viruses are mainly due to vertical transmission. These viruses also suggest possible interaction between mosquito and different types of hosts for blood meal as viruses from birds, amphibians and mammals (e.g. *Porcine mastadenovirus C)* were most commonly found. Few studies have also reported that mosquitoes feed on nectar and plant sap for intake of carbohydrates as energy source, and this is attributed to presence of wide variety of plant viruses (e.g. *Tospovirus*, *Cucurbit chlorotic yellows virus*, *Grapevine geminivirus A)* in virome[27]. A high proportion of reads were unclassified (50%), which shows limitation of incomplete or insufficient viral reference databases, sequencing error or presence of novel or extreme divergent viral sequences. The data is restricted to the workflow chosen for the analysis in Commander. Hence, limitations suggest the importance of expansion of viral reference databases, improving workflows and bioinformatic tools specifically for nanopore sequencing data to incorporate more insightful information.

The total viral profile observed in this study was quite stable and aligns with the previous studies highlighting mosquito virome composition with majorly dominated ISVs and little or no presence of pathogenic arboviruses. This study with limited samples size and sequencing depth was able to provide baseline data with viral profiling for *Aedes* mosquito population circulating in central India.

Mosquito could be a useful substitute sample for routine arbovirus surveillance or molecular xeno-monitoring incorporating NGS technology. This has the potential to substitute human sampling for surveillance of arboviruses[36]. In situations where mass drug administration and elimination strategy are thought of as a disease control, the proxy also aids in identifying the effects of mass therapy and in confirming the eradication of neglected mosquito-borne illnesses. This would allow the objective identification of arboviruses that are significant for public health from a single sample, thereby expanding the scope of surveillance. Moreover, NGS applications will probably lead to the development of new, more advanced protocols, which will raise the

platform's sensitivity and efficiency. Effective applications of the portable MinION device for disease surveillance have been documented during the Ebola outbreak in West Africa. The real-time functionality is essential for this goal since it allows for the quick establishment of progression patterns through the monitoring of disease spread. The development of effective algorithms for the Oxford Nanopore Technology data and relatively higher error rates presents a significant challenge[37].

Overall, this study reveals the effective approach of using unbiased metagenomic pipeline for characterization of viral diversity in mosquitoes without prior viral enrichment. While detection of viral sequences does not primarily indicate active infection but it provides a comprehensive idea for further investigations. This approach can be used further for large number of field-collected mosquitoes that will contribute to the preparedness against emerging or re-emerging viruses or novel viruses for early warning against future outbreaks. This approach also has the potential to be used for regular monitoring of endemic viruses in a particular location or a season.

Data availability

The raw sequence reads generated in this study are available at the NCBI Sequence Read Archive (SRA) database under BioProject PRJNA1369453 (https://www.ncbi.nlm.nih.gov/sra/PRJNA1369453). Biosamples SAMN53389395, SAMN53389396, SAMN53389397, SAMN53389398, SAMN53389399, SAMN53389400, SAMN53389401, SAMN53389402, SAMN53389403, SAMN53389404. The assembled viral reads generated from Genome Detective platform have been deposited in GenBank under accession numbers PZ192165 - PZ192403.

Received: 28 October 2025; Accepted: 13 April 2026
Published online: 25 April 2026

References

1. World Health Organisation (WHO). Accessed on 06 Aug (2024). Available on: https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases.
2. Martin, E. et al. Mosquito-borne viruses and insect-specific viruses revealed in field-collected mosquitoes by a monitoring tool adapted from a microbial detection array. *Appl. Environ. Microbiol.* **85** (19), e01202–e01219 (2019).
3. Pan, Y. F. et al. Metagenomic analysis of individual mosquito viromes reveals the geographical patterns and drivers of viral diversity. *Nat. Ecol. Evol.* **8** (5), 947–959 (2024).
4. National centre for vector borne disease control (NCVBDC). Accessed on 06 Aug 2024. Available on: https://ncvbdc.mohfw.gov.in/index.php.
5. Kuno, G. Universal diagnostic RT-PCR protocol for arboviruses. *J. Virol. Methods* **72**(1), 27–41 (1998).
6. Lanciotti, R. S. et al. Rapid detection of West Nile virus from human clinical specimens, field-collected mosquitoes, and avian samples by a TaqMan reverse transcriptase-PCR assay. *J. Clin. Microbiol.* **38**(11), 4066–4071 (2000).
7. Scaramozzino, N. et al. Comparison of flavivirus universal primer pairs and development of a rapid, highly sensitive heminested reverse transcription-PCR assay for detection of flaviviruses targeted to a conserved region of the NS5 gene sequences. *J. Clin. Microbiol.* **39**(5), 1922–1927 (2001).
8. Eshoo, M. W. et al. Direct broad-range detection of alphaviruses in mosquito extracts. *Virology* **368**(2), 286–295 (2007).
9. Osei-Poku, J., Mbogo, C. M., Palmer, W. J. & Jiggins, F. M. Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Mol. Ecol.* **21**(20), 5138–5150 (2012).
10. Chandler, J. A., Liu, R. M. & Bennett, S. N. RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. *Front. Microbiol.* **6**, 185 (2015).
11. Kumar, A., Murthy, S. & Kapoor, A. Evolution of selective-sequencing approaches for virus discovery and virome analysis. *Virus Res.* **239**, 172–179 (2017).
12. Russell, J. A. et al. Unbiased strain-typing of arbovirus directly from mosquitoes using nanopore sequencing: a field-forward biosurveillance protocol. *Sci. Rep.* **8**, 5417 (2018).
13. Gangopadhayya, A. et al. Metagenomic analysis of viromes of *Aedes* mosquitoes across India. *Viruses* **16**(1), 109 (2024).
14. Hameed, M. et al. A metagenomic analysis of mosquito virome collected from different animal farms at Yunnan–Myanmar Border of China. *Front. Microbiol.* **11**, 591478 (2021).
15. Shi, C. et al. Stable distinct core eukaryotic viromes in different mosquito species from Guadeloupe, using single mosquito viral metagenomics. *Microbiome* **7**, 1–20 (2019).
16. Tyagi, B. K., Munirathinam, A. & Venkatesh, A. A catalogue of Indian mosquitoes. *Int. J. Mosq. Res.* **2**(2), 50–97 (2015).
17. Sarma, D. K. et al. Molecular surveillance of dengue virus in field-collected *Aedes* mosquitoes from Bhopal, central India: Evidence of circulation of a new lineage of serotype 2. *Front. Microbiol.* **14**, 1260812 (2023).
18. Agarwal, A. et al. Application of real-time RT-PCR in vector surveillance and assessment of replication kinetics of an emerging novel ECSA genotype of Chikungunya virus in Aedes aegypti. *J. Virol. Methods*. **193**(2), 419–425 (2013).
19. Hameed, M. et al. A viral metagenomic analysis reveals rich viral abundance and diversity in mosquitoes from pig farms. *Transbound. Emerg. Dis.* **67**(1), 328–343 (2020).
20. Karamitros, T. & Magiorkinis, G. Multiplexed targeted sequencing for oxford nanopore MinION: A detailed library preparation procedure. in *Next Generation Sequencing: Methods Protocols* 43–51 (2018).
21. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford nanopore sequencing. *Genome Biol.* **20**, 1–0 (2019).
22. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**(4), 1160–1166 (2019).
23. Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F. & Reimer, A. Metagenomics: the next culture-independent game changer. *Front. Microbiol.* **8**, 1069 (2017).
24. Zhang, Y. Z., Chen, Y. M., Wang, W., Qin, X. C. & Holmes, E. C. Expanding the RNA virosphere by unbiased metagenomics. *Annual Rev. Virol.* **6**, 119–139 (2019).
25. Greninger, A. L. et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* **7**, 1–3 (2015).
26. Ergunay, K. et al. Impact of nanopore-based metagenome sequencing on tick-borne virus detection. *Front. Microbiol.* **14**, 1177651 (2023).
27. Ali, R. et al. Characterization of the virome associated with Haemagogus mosquitoes in Trinidad, West Indies. *Sci. Rep.* **11**(1), 16584 (2021).
28. Zakrzewski, M. et al. Mapping the virome in wild-caught Aedes aegypti from Cairns and Bangkok. *Sci. Rep.* **8**(1), 4690 (2018).

29. Olmo, R. P. et al. Mosquito vector competence for dengue is modulated by insect-specific viruses. *Nat. Microbiol.* **8**, 135–149. https://doi.org/10.1038/s41564-022-01289-4 (2023).
30. Zhou, J. et al. Complete genomic and ultrastructural analysis of a Nam Dinh virus isolated from *Culex pipiens quinquefasciatus* in China. *Sci. Rep.* **7**(1), 271 (2017).
31. Shi, M. et al. Redefining the invertebrate RNA virosphere. *Nature* **540**(7634), 539–543 (2016).
32. Xu, Y., Xu, J., Liu, T., Liu, P. & Chen, X. G. Metagenomic analysis reveals the virome profiles of *Aedes albopictus* in Guangzhou, China. *Front. Cell. Infect. Microbiol.* **13**, 1133120 (2023).
33. Baidaliuk, A. et al. Cell-fusing agent virus reduces arbovirus dissemination in *Aedes aegypti* mosquitoes in vivo. *J. Virol.* **93**(18), 10–128 (2019).
34. Bolling, B. G., Olea-Popelka, F. J., Eisen, L., Moore, C. G. & Blair, C. D. Transmission dynamics of an insect-specific flavivirus in a naturally infected Culex pipiens laboratory colony and effects of co-infection on vector competence for West Nile virus. *Virology* **427**(2), 90–97 (2012).
35. Carvalho, V. L. & Long, M. T. Insect-specific viruses: an overview and their relationship to arboviruses of concern to humans and animals. *Virology* **557**, 34–43 (2021).
36. Geleta, D. & Yewhalaw, D. Molecular xenomonitoring for surveillance of mosquito-borne diseases.
37. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION sequencing and genome assembly. *Genom. Proteom. Bioinform.* **14**(5), 265–267 (2016).

Acknowledgements

I am grateful for the assistance by the District Malaria Officer, Gwalior in collection of mosquito larvae. I, sincerely, thank the Director, Defence Research and Development Establishment (DRDE), Ministry of Defence, Govt. of India for their endless support and encouragement to carry out this work. I thank CSIR-UGC for the Senior Research Fellowship. The DRDE accession number for this manuscript is DRDE-IREC-187-28/10/2024.

Author contributions

Ekta Gupta collected samples, performed experiments and wrote manuscript under supervision of Dr. Shashi Sharma and Dr. Paban Kumar Dash. Dr. Shashi Sharma performed large data analysis and provided NGS facility . Dr. Paban Kumar Dash and Dr. Manmohan Parida designed study and reviewed manuscript. All authors provided substantial input and approved manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

This study was performed after seeking approval by DRDE Institutional Biosafety Committee vide the no. IBSC/VIRO02/2020/PKD. The experimental protocols were executed under the study according to all proper biosafety and the regulatory guidelines.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-49112-y.

Correspondence and requests for materials should be addressed to P.K.D.